



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Recovering discourse relations: Varying influence of discourse adverbials

Citation for published version:

Rohde, H, Dickinson, A, Clark, C, Louis, A & Webber, B 2015, Recovering discourse relations: Varying influence of discourse adverbials. in *Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics*. <<http://homepages.inf.ed.ac.uk/mroth/LSDSem/#program>>

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Recovering discourse relations: Varying influence of discourse adverbials

Hannah Rohde Anna Dickinson Chris Clark Annie Louis Bonnie Webber

University of Edinburgh

[hannah.rohde, bonnie.webber]@ed.ac.uk

[anna.y.dickinson, chrisclark272, annieplouis]@gmail.com

Abstract

Discourse relations are a bridge between sentence-level semantics and discourse-level semantics. They can be signalled explicitly with discourse connectives or conveyed implicitly, to be inferred by a comprehender. The same discourse units can be related in more than one way, signalled by multiple connectives. But multiple connectives aren't necessary: Multiple relations can be conveyed even when only one connective is explicit. This paper describes the initial phase in a larger experimental study aimed at answering two questions: (1) Given an explicit discourse adverbial, what discourse relation(s) do naive subjects take to be operative, and (2) Can this be predicted on the basis of the explicit adverbial alone, or does it depend instead on other factors?

1 Introduction

Semantics comes both explicitly and implicitly from a text. One bridge between sentence-level semantics and discourse semantics consists of relations between sentences and/or clauses, called variously *discourse relations* (Prasad et al., 2014), *coherence relations* (Kehler, 2002) or *rhetorical relations* (Mann and Thompson, 1988). Such relations between what we will call here *discourse spans* can be signaled explicitly via discourse connectives or specific lexico-syntactic constructions, or conveyed implicitly, via inference on the part of a comprehender. But **when** does the latter happen? Previously, it was assumed that relations are conveyed implicitly when they are **not** signalled explicitly. But consider Ex. 1a-b, each with two explicit connectives conveying distinct relations:

- (1) a. Let's eat dinner now because otherwise we'll miss the film.

- b. I can't walk 5 miles, so instead I'll take a taxi.

In Ex. 1a, *because* signals the REASON for eating dinner now, while *otherwise* signals the CONDITION under which we'll miss the film. In Ex. 1b, *so* signals the RESULT of my inability to walk so far, while *instead* signals the CHOSEN ALTERNATIVE to taking a taxi.¹

However, both relations may still be conveyed, even if only one is signalled explicitly, as in Ex. 2a-c:

- (2) a. Let's eat dinner now. Otherwise we'll miss the film.
b. I can't walk 5 miles. Instead I'll take a taxi.
c. I can't walk 5 miles, so I'll take a taxi.
d. Let's eat dinner now because we'll miss the film.

So it is not the case that implicit discourse relations only arise when discourse relations are **not** signalled explicitly. (Ex. 2d shows that a CHOSEN ALTERNATIVE is not achieved with the single connective *because*.)

The potential availability of multiple concurrent discourse relations raises important questions for both Language Technology (LT) and psycholinguistics: When a discourse relation is signalled with an explicit connective, should a LT system also look for a distinct implicit relation? From the perspective of psycholinguistics, implicit co-occurring relations raise fundamental questions about how comprehenders infer discourse relations and which contexts allow such relations to be understood without an explicit linguistic signal.

¹The sense labels used here (in small caps) are short forms of the labels used in the PDTB 2.0 (Prasad et al., 2008; Prasad et al., 2014).

Despite multiple explicit connectives being observed in Catalan and Spanish (Cuenca and Marin, 2009) as well as Turkish (Zeyrek, 2014) and English, questions about multiple relations in the presence of only a single connective have not yet been addressed (Section 2). To address them, we have embarked on a large crowd-sourcing experiment, the first phase of which is described in Sections 3–5. Section 6 discusses our results to date, with further phases described in Section 7.

2 Background

This is not the first work to call attention to multiple co-occurring connectives. Webber and colleagues (1999) used them to argue that discourse spans could be related by both adjacency relations and anaphoric relations. Similarly, in the context of Catalan and Spanish oral narrative, Cuenca and Marin (2009) used them to argue for different patterns and degrees of discourse cohesion. Oates (2000) considers how multiple discourse connectives should be used in Natural Language Generation, noting that the order in which they occur correlates with the hierarchy of discourse connectives presented in (Knott, 1996). Fraser (2013) considers the order in which multiple CONTRASTIVE connectives co-occur, describing their patterning in terms of *general contrastive* discourse markers and *specific contrastive* discourse markers. For Turkish, Zeyrek (2014) describes patterns of multiple co-occurring connectives that signal CONTRASTIVE and/or CONCESSIVE relations.

These efforts have all been directed at providing an account of the existence of multiple connectives and their patterning. As for the phenomenon illustrated in Ex. 2a–c, the only work we are aware of is an MSc project supervised by the first co-author (Rohde). This study, by Xi Jiang (2013), involved four discourse adverbials (*after all*, *instead*, *in fact*, *in general*) that can occur alone or following a conjunction. Jiang presented participants in a crowd-sourcing experiment with a set of fill-in-the-gap passages such as the following

- (3) Logically, she should be dead / _____ instead / she feels fine, caring for her daughters and walking a pedometer-measured two miles a day.
- (4) He suspected he shouldn't say that / _____ instead / he lied.

asking the participants to either insert one of five conjunctions (*and*, *because*, *but*, *or*, *so*) into the

gap or choose *None*.² In half the passages (10 per adverbial), the author had used one of these conjunctions before the adverbial (which Jiang then removed), and in the other half (including Ex. 3–4), the author had used no conjunction before the adverbial. The only criteria used in selecting these passages were brevity (i.e., could the passage be read quickly?) and clarity (i.e., did the passage make sense when presented out of context?).

Jiang's study was aimed at answering two questions: (1) When the author had used an explicit conjunction before the discourse adverbial, did participants always fill the gap with the same conjunction; and (2) where the original passage lacked an explicit conjunction, did participants choose to omit an explicit conjunction (i.e., did they choose *None*).

Each of the 80 passages (20 per adverbial) was annotated by the same 52 participants. Jiang's results showed some interesting patterns. In the gap preceding *after all*, participants tended to insert *because*, indicating that they interpreted the content of the second span as a REASON for the content of the first span, independent of whether the original text contained *because* or a different conjunction or *None*. In contrast, in the gap preceding *instead*, the choice made by participants varied from passage to passage: For instance, they reliably inserted *but* in Ex. 3 and *so* in Ex. 4, even though the original text contained no conjunction.

The data that Jiang collected suggested that the answer to both of her questions was **no**, but stopped there. One reason is that the response *None* was ambiguous: Participants could have used it to mean "I can't insert any of these conjunctions to express the sense I get", or "The sense I get cannot be expressed with a conjunction", or "I don't get any additional sense". Secondly, in using only brevity and clarity as her criteria for selecting passages from COCA, Jiang did not assess whether all the conjunction-less passages she selected might have been similar in terms of how their clauses/sentences related and hence would all draw the same response from participants. Finally, Jiang only considered four adverbs, so could not draw more general conclusions. The current, much larger enterprise attempts to avoid these problems.

²All passages were taken from the Corpus of Contemporary American English (COCA) corpus.byu.edu/coca/.

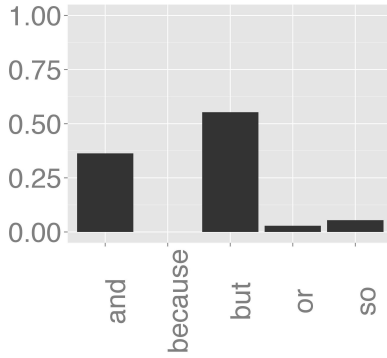


Figure 1: The distribution of $\langle \text{conjunction} \rangle$ immediately preceding *instead* tokens in Google NGRAMS, with or without a comma after the conjunction, excluding cases where *instead* was followed by *of*.

3 Task Definition

We have embarked on a large-scale study of discourse adverbials, attempting to gather evidence that will help answer two specific questions:

1. Given an explicit discourse adverbial in a passage, what discourse relation(s) do naive subjects take to be operative?
2. Can the relation(s) be predicted on the basis of the explicit adverbial alone, or does it depend on the arguments to the relation or on everything in the passage?

Note that the discourse relations that subjects take to be operative may corroborate the sense conveyed by the discourse adverbial or they may be distinct.

In this paper we describe Phase I of the study, carried out between August 2014 and June 2015. We began with a survey of Google NGRAMs to first establish the overall frequency and preferred conjunction(s) of a wide range of adverbials. In the long term, our study aims to examine both common and rarer adverbials (see Section 7) and those with a single preferred co-occurring conjunction and those with a flatter distribution. As Figures 1 and 2 show, the distribution of conjunctions is neither uniform for a given adverbial nor equivalent across adverbials. Since all four adverbials (*after all*, *instead*, *in general* and *in fact*) used in (Jiang, 2013) had different distributions, we decided to target the same adverbials in our Phase I study.

Also following (Jiang, 2013), we wanted to see whether subjects responded differently to pas-

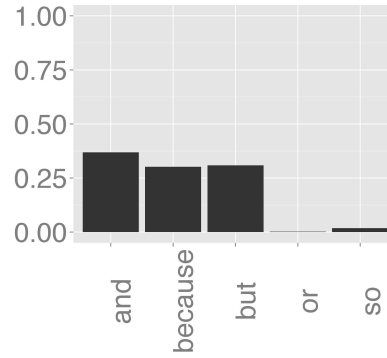


Figure 2: The distribution of $\langle \text{conjunction} \rangle$ immediately preceding *after all* tokens in Google NGRAMS.

sages in which the author explicitly used a pair of connectives (i.e. $\langle \text{conjunction} \rangle$ – $\langle \text{adverbial} \rangle$) compared with those in which the author only used an explicit adverbial. The former we call *explicit passages* and the latter, *implicit passages*.

4 Phase I Experiment

Each participant in Phase 1 saw 50 passages, each containing a gap between two spans of text, the second beginning with a discourse adverbial, as in Ex. 3–4. With *explicit passages*, we replaced the conjunction with a gap, while with *implicit passages*, we inserted a gap before the adverbial. For each passage, participants were instructed to fill in the gap with the word of their choice (from a randomly ordered list of the six conjunctions *and*, *because*, *before*, *but*, *or*, *so*) that “best reflects the **meaning** of the connection” between the spans. They also had the option of choosing either *None at all* (for cases where they felt that no conjunction was possible) or *Other word or phrase* (for cases where they felt that only some option other than the six presented conjunctions was appropriate). These were intended to correct for the ambiguity of *None* in Jiang’s study.

At a coarse sense level, all six conjunctions are relatively unambiguous: Table 1 shows the frequency of their main sense in the Penn Discourse TreeBank (The PDTB Research Group, 2008). As such, there are grounds for believing that the experiment targeted the participants’ inferred relation through choosing a conjunction that realizes it, even if the sense is only a coarse one.

	# tokens	sense label	overall %
<i>and</i>	3000	CONJUNCTION	91.0%
<i>because</i>	858	REASON	99.5%
<i>before</i>	326	PRECEDENCE	99.0%
<i>but</i>	3308	COMPARISON ³	90.7%
<i>or</i>	98	ALTERNATIVE	86.7%
<i>so</i>	263	RESULT	99.6%

Table 1: Proportion of explicit tokens of each conjunction having its most frequent sense label

4.1 Interface

Working with a group of researchers and a pilot group of participants, we iteratively designed and evaluated an interface and a set of instructions aimed at encouraging participants to choose a conjunction that identified the sense of the connection between the two spans of text in a passage — the span before the gap and the span following it. Instructions for the task could be reviewed when necessary by clicking on a button labelled “Show Instructions”, to the right of the heading “Trial” (Figure 3).

During pilot testing, it emerged that participants sometimes chose *None at all* when it sounded more fluent and less awkward to them than did an explicit conjunction. In order to avoid this, we explicitly instructed participants to choose the conjunction that best conveyed the sense of the connection, “even if the resulting text sounds awkward”, but also offered them the opportunity to record whether they would in fact use the chosen conjunction, or whether it sounded odd to them in that context (Figure 4).

To avoid order effects, the stimuli were pseudo-randomised for each participant such that each participant only saw each excerpt once, they never encountered more than three of the same adverbial in a row, and for explicit passages, they never saw excerpts expecting the same conjunction more than three times in a row. In addition, the list of conjunctions appeared in a different order for each participant, to avoid the risk of skewing the results, should participants prefer conjunctions presented at the top of the list.

After a participant had read the instructions, three practice items were presented.

4.2 Stimuli

Of the 50 passages used in Phase 1, 38 replicated those previously used in (Jiang, 2013). Of the remaining twelve, eight came from a large set of possible stimuli we collected from the New

York Times Annotated Corpus (NYTAC) (Sandhaus, 2008) for use in later phases of the experiment, while four were “catch trials”, intended to ensure participants were paying attention. Table 2 shows the number of *explicit passages* for each of the four adverbials (where the explicit conjunction before the adverbial was deleted, leaving a gap) and *implicit passages* (where a gap was simply inserted before the adverbial).

	explicit	implicit
<i>after all</i>	7	5
<i>in fact</i>	7	4
<i>in general</i>	7	5
<i>instead</i>	6	5

Table 2: Explicit/implicit passages per adverbial

The 38 excerpts from Jiang were chosen based on the responses they had received during her study. For example, for the *instead* implicits, two showed a range of responses, one showed participant agreement on *but*, one showed agreement on *because*, and one showed agreement on *so*. The eight new stimuli from NYTAC (two per adverbial) were longer and more complex than those used in (Jiang, 2013). The purpose of these stimuli, besides providing more data, was to identify participants who were discouraged or confused by these passages, since later phases of the experiment would use stimuli drawn only from NYTAC.

4.3 Participants

Seventy participants, all with addresses in the United States, completed the trial through Amazon Mechanical Turk. Demographic data collected in a short questionnaire before the main trial showed that participants were aged 20-67 (mean 37), 71% read newspapers at least twice a week, and half were female. All were English speakers. Each participant was paid \$8 for their contribution.

5 Phase 1 Results

All participants paid attention to the task, as indicated by their selection of sensible responses for the four catch trials, while they varied in how long the task took them and how often they agreed with the choices made by other participants. As we required fewer participants to complete Phase 2 of the task, we reduced the participant number based on their performance in Phase 1. Specifically, we removed data from 12 participants with very short completion rates and high rates of disagreement

ConnTextUniversity of Edinburgh

TrialShow Instructions

I don't mind **walking** // **in fact** it's good exercise

Conjunction:

☐ Or

☐ But

☐ Because

☐ None at all

☐ So

☐ And

☐ Before

☐ Other word or phrase

Once you have made your selections, press submit to complete the trial. To share additional comments about this trial, please [click here](#).

Submit

Figure 3: Screen shot of passage presentation

ConnTextUniversity of Edinburgh

TrialShow Instructions

Their mother has lived in the United States longer than she lived in the Philippines. Her brother, who lives in Manila, scolds her: "We all thought you would become a **nun** // **but instead** you became an American!"

Conjunction:

☐ Or

☒ But

☐ Because

☐ None at all

☐ So

☐ And

☐ Before

☐ Other word or phrase

Does 'but' sound okay?

☐ I could say it this way

☐ It sounds strange here

Once you have made your selections, press submit to complete the trial. To share additional comments about this trial, please [click here](#).

Submit

Figure 4: Screen shot of a participant being asked to indicate whether or not their choice of a conjunction that fits with respect to its sense — in this case, “but” — sounds natural

with other participants, as well as 3 trials in which a participant selected the response *before*, which was intended for use in only the catch trials. The resulting dataset of responses from 58 participants comprises 2665 judgments over the 46 target passages (ignoring the four catch trials). The results reported below are raw counts, and do not yet take account of potential participant bias (Passeau and Carpenter, 2014).

Considering the dataset as a whole, we can ask how often a participant’s response matched the author’s original choice. (Note that this can only be assessed on *explicit passages* – that is, ones where the author explicitly used a pair of co-occurring connectives, cf. Section 3). Table 3 shows the pattern of participant responses for passages for which the authors themselves had included an explicit conjunction before the adverbial. Recall that participants always saw a gap before the discourse adverbial, regardless of the author’s original choice to use or not use a conjunction, meaning the explicit and implicit passages were indistinguishable.

	AND	BECAUSE	BUT	OR	SO
and	189	14	81	5	33
because	60	105	60	2	9
but	68	48	497	7	9
or	2	0	2	35	0
so	125	1	25	2	56
other	3	1	8	2	0
none at all	17	4	23	5	9

Table 3: Confusion matrix for explicit passages. Rows show participant responses (participants’ selected conjunctions in lower case) for passages whose authors had used the explicit conjunctions in the columns (in CAPS).

The values on the diagonal in Table 3 show a high degree of convergence between participant and author choices: The largest value for any column and any row is the value indicating participant~author agreement. A conjunction like *and* notoriously underspecifies the relation sense since it is compatible with many senses. The results in Table 3 allow us to ask what more specific senses participants infer in cases in which the original author used *and*. Although participants overall favor ‘and’ for author ‘AND’ (189 instances out of 464 ‘AND’ trials), they also show a preference for the inference of causality with their selection of *so* (125 instances out of the 464 ‘AND’ trials).

Table 4 shows the pattern of responses for pas-

sages in which the author did not include a conjunction. In only a small number of cases (69 instances out of 1158 ‘NONE’ trials) did a participant choose *None at all*. Therefore, in answer to question (1) from Section 3, participants are able to reliably select an explicit conjunction that realizes the relation(s) they take to be operative. The next section considers participants’ behavior for each of the 4 adverbials in turn.

	author=NONE
and	275
because	404
but	252
or	1
so	147
other	10
none	69

Table 4: Response distribution (implicit passages)

5.1 Variation across adverbials for explicit passages

To address the second question raised in Section 3, we analyzed participant responses to each adverbial. Tables 5-7 show the responses for *after all*, *in fact*, *in general* and *instead* respectively, when the original author included an explicit conjunction.

	AND	BECAUSE	BUT
and	18	6	30
because	9	51	51
but	25	0	128
or	0	0	0
so	0	0	3
other	1	0	3
none	5	1	17

Table 5: Explicit *after all* response distribution. Participant responses in lower case, versus author choice in CAPS. (Six explicit *after all* passages — 1 AND, 1 BECAUSE, 4 BUT)

For *after all* (Table 5), participants assigned a meaning of *because* not only for author BECAUSE but frequently for author BUT and AND. This is particularly odd for BECAUSE and BUT, since however underspecified one might take these two conjunctions to be, the senses they convey are still different. This suggests that the adverbial itself may be biasing the inferred relation.

For *in fact* (Table 6), the responses track the authors with two notable exceptions. First, the responses show that author BUT and author SO passages are frequently labelled by participants with

	AND	BECAUSE	BUT	OR	SO
and	53	8	27	5	29
because	1	54	4	2	1
but	1	48	74	7	6
or	0	0	0	35	0
so	0	1	4	2	15
other	0	1	5	2	0
none	3	3	2	5	7

Table 6: Explicit *in fact* response distribution. Participant responses in lower case versus author choice in CAPS. (Seven explicit *in fact* passages — 1 AND, 2 BECAUSE, 2 BUT, 1 OR, 1 SO)

the less specific conjunction *and*. This may reflect the fact that the conjunction that most frequently appears left-adjacent to *in fact* is *and*, according to our study of the Google NGRAM corpus (cf. Section 3). Second, cases of author BECAUSE are split closely between responses of *because* and *but*. The alternation between “because” and “but” responses is surprising (as already noted above), given that they are not typically understood to be synonyms or even hyponyms or hypernyms. Nor does this variation appear to simply reflect a scenario in which, of the two BECAUSE passages, one favored *because* while the other favored *but*: Rather, each BECAUSE passage (such as Ex. 5) received a mix of *because* and *but* responses.

- (5) Americans’ big-is-better mentality is a shame in the case of artichokes _____ in fact the small ones are much easier to clean, cook more quickly and can be purchased spontaneously because they don’t take any more time than any other vegetables.

	AND	BUT	SO
and	102	23	4
because	50	4	8
but	36	85	3
or	2	0	0
so	33	1	41
other	2	0	0
none	7	3	2

Table 7: Explicit *in general* response distribution. Participant responses in lower case versus author choice in CAPS. (Seven *in general* passages — 4 AND, 2 BUT, 1 SO)

For *in general* (Table 7), the responses track the author, suggesting that the adverbial itself is not biasing the inferred relation, but that responses depend on properties of the adjacent clauses or the larger context.

	AND	BUT
and	16	1
because	0	1
but	6	210
or	0	2
so	92	17
other	0	0
none	2	1

Table 8: Explicit *instead* response distribution. Participant responses in lower case versus author choice in CAPS (Six *instead* passages — 4 AND, 2 BUT)⁵

Like the data for *in fact*, the data for *instead* (Table 8) highlight a link between *and* and *so*, but in the opposite direction. For *in fact*, author SO received many *and* responses, whereas for *instead*, it was the reverse: Author AND received many *so* responses. This is in keeping with the observation that *and* is underspecified but is compatible with, and often implicates, a temporal or causal relationship between the eventualities denoted by the adjacent clauses (Gazdar, 1979). With *in fact*, participants are selecting a less specific conjunction (*and*) rather than the more specific *but* or *so*, whereas for *instead*, they are selecting the more specific *so*. It is possible that this can be explained as a frequency-induced bias: Compared with *in fact*, our Google NGRAM estimates show *instead* to have proportionally more co-occurrences with *so*, potentially leading participants to posit “so instead” for passages whose author had used AND.

5.2 Variation across adverbials for implicit passages

As noted earlier, Jiang’s (2013) study leaves open the question of how to interpret a *None* response from a participant: Does it mean the participant believed there was no relation to infer, or that none of the available conjunctions were appropriate, or that there was an inferred relation but the resulting passage simply sounded awkward? Our experiment was designed to eliminate this ambiguity. That is, *None* can be understood to convey “no relation to infer”, given that participants could choose *Other* if they wanted to fill in an alternative conjunction or they could mark the meaning they inferred but then tag it as awkward with the “would not say” button. Note that in Jiang’s study, 15.7% of the responses were *None*. In our study, the proportion was comparable, with 15.2% of responses being one of our variants of *None*, i.e.

None (7.7%), *Other* (1.0%) or marked as something the participant would not say (6.4% of responses).

Our data on implicit passages therefore provides a clearer picture of how frequently participants assign a conjunction even when the author had used no conjunction. The results in Table 9 show that no adverbial favors *None* in these cases: *after all* had only 26/348 judgments; *in fact*, only 20/231; *in general*, only 13/290, and *instead*, only 10/290.

	after all	in fact	in general	instead
and	50	87	118	20
because	245	35	86	38
but	16	83	50	103
or	1	0	0	0
so	4	3	21	119
other	5	3	2	0
none	26	20	13	10

Table 9: Response distribution for implicit passages by adverbial (20 unique passages: 6 “after all”, 4 “in fact”, 5 “in general”, 5 “instead”)

Table 9 also confirms some of the behavior observed in the responses to explicit passages. First, *after all* shows a preference for the response *because*, whereas *in fact*, *in general* and *instead* all show more variability. This variability suggests that participants are responding to the content of the conjoined arguments to identify the sense, rather than associating the adverbial with one preferred connective. According to our Google NGRAM estimates, *after all* differs from the other three adverbials insofar as *because* is one of its most frequent co-occurring conjunctions. In contrast, *in fact*, *in general* and *instead* rarely co-occur with *because*. So participant behavior may reflect their sensitivity to the affinity of *after all* for *because*.

Finally, we can check how consistent participants were in selecting their response to each implicit passage. For each passage, we identify the most frequent response and the proportion of participants who selected that response. For all passages, the most frequent response was neither *None* nor *Other*. Table 10 shows the mean agreement for each adverbial, collapsed across passages, revealing whether different adverbials demonstrate different degrees of inter-annotator consistency. Table 10 shows that the agreement rate for two adverbials (*after all* and *instead*) is higher than for the other two: *After all* consistently

avored *because*, while *instead* showed more variability in inferred conjunctions but nevertheless had a similar agreement rate. So while the four adverbials have different degrees of overall inter-annotator consistency on *implicit passages*, none of them shows random selection over the five non-None/non-Other responses, which would yield an agreement rate of just over 0.2.

after all	in fact	in general	instead
0.706	0.581	0.503	0.717

Table 10: Participant agreement rates by adverbial

6 Discussion

We draw two conclusions from Phase 1 of our study: (1) It is possible for naive subjects to infer an implicit conjunction alongside an explicit discourse adverbial, even for passages in which the original author used only an explicit adverbial, and (2) subjects do so reliably and systematically, depending on the adverb. Our subjects had the option on each trial to decline to add a conjunction, but they did not. Rather, they endorsed meaning-bearing conjunctions and did so in a way that is not explainable from the adverbial alone. In other words, it is not the case that any of these four adverbials is uniformly associated with a single conjunction whose meaning is linked directly to that of the adverbial itself. That would not explain the fact that, across passages, different conjunctions were endorsed as plausible insertions for the same adverbial. What’s more, the selection of a conjunction for a given passage shows a strong degree of consistency, particularly for *after all* and *instead*.

The second point is that discourse adverbials themselves are not indiscriminate with regard to the conjunction that they appear to favor. The analysis of *after all* showed that participants selected a causal interpretation (*because*) more often than would be expected based on the conjunction provided by the original author and with a bias that was more pronounced than in passages with any other adverbial. This highlights potential differences among adverbials (either individually or by class): Not all adverbials may be compatible with all conjunctions. Even where variation is permitted, the adverbial may bring its own preference to bear on the inference of an additional co-occurring relation. This point was made by Jiang (2013) as

well, and our data are in keeping with the range of behavior she reports across these four adverbials. The new study goes beyond prior work by ensuring that participants who preferred to communicate that none of the available conjunctions should be inserted had recourse to three distinct responses: a stylistic rejection of the selected conjunction (“does it sound okay?”), an option to insert an alternative conjunction (*Other*), or simply the response *None at all* (to reject insertion of any explicit connective to link the two spans of text).

So how do participants identify the conjunction they insert into these passages? One hypothesis might be that the purported lexical semantics of the adverbial is what determines its co-occurrences with conjunctions. Under that account, *instead* might be expected to favor a conjunction that expresses contrast, i.e., *but*. The distribution of responses for explicit passages with *instead* shows that *but* was indeed the preferred response when the author chose to use *but*. However, when the author used *and*, participants favored *so*, which generally conveys RESULT. For implicit passages with *instead*, the response choice *but* was likewise frequent, but not as frequent as *so*. On the other hand, the results for *after all* do suggest that the inference of *because* is common when that adverbial is present. This pattern is there for the explicit passages, and is even more evident for the implicit passages (for which 245 of 348 responses were *because*). This finding could suggest that *after all* either conveys a single sense itself or is used frequently in contexts in which a REASON relation is operative. The other adverbials show no such preference, implying that it is properties of the clauses themselves and the rest of the discourse that allow a consistent meaning to be identified for each passage.

7 Future work

Building on the results of Phase 1, we have begun to run a larger Phase 2 study with twenty adverbials, using 976 excerpts. The 58 participants whose results are reported here for Phase 1, have been invited to complete further Amazon Turk hits. In the longer term, we hope to explore the other common case of non-adjacent co-occurring discourse connectives, as in

- (6) They cut few trees in the summer, when they prefer to feed more on fresh grasses, tubers, and saplings, **but** au-

tumn, **however**, is a period of intensive logging for beavers. (hawriver.org/peaceful-coexistence-with-beavers)

and to extend the research cross-linguistically.

Acknowledgments

This research has been supported by the Nuance Foundation. The authors would also like to acknowledge advice and support from researchers on the Penn Discourse TreeBank and helpful comments from our three anonymous reviewers.

References

- M.J. Cuenca and M.J. Marin. 2009. Co-occurrence of discourse markers in Catalan and Spanish oral narrative. *Journal of Pragmatics*, 41(5):899–914.
- Bruce Fraser. 2013. Combinations of contrastive discourse markers in english. *International Review of Pragmatics*, 5:318–340.
- G. Gazdar. 1979. *Pragmatics: Implicature, Presupposition, and Logical Form*. Academic Press, London.
- Xi Jiang. 2013. Predicting the use and interpretation of implicit and explicit discourse connectives. Master’s thesis, Linguistics and English Language, University of Edinburgh. MSc in English Language.
- Andrew Kehler. 2002. *Coherence, Reference, and the Theory of Grammar*. CSLI Publications, Stanford.
- Alistair Knott. 1996. *A Data-driven Methodology for Motivating a Set of Coherence Relations*. Ph.D. thesis, Department of Artificial Intelligence, University of Edinburgh.
- William C. Mann and Sandra A. Thompson. 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.
- Sarah Oates. 2000. Multiple discourse marker occurrence: Creating hierarchies for natural language generation. In *Proceedings, ANLP-NAACL 2000 Student Research Workshop*, pages 41–45, Seattle WA.
- Rebecca Passonneau and Bob Carpenter. 2014. The benefits of a model of annotation. *Transactions of the Association of Computational Linguistics*, 2(1):311–326.
- Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltasakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings, 6th International Conference on Language Resources and Evaluation*, pages 2961–2968, Marrakech, Morocco.

- Rashmi Prasad, Bonnie Webber, and Aravind Joshi. 2014. Reflections on the penn discourse treebank, comparable corpora and complementary annotation. *Computational Linguistics*, 40(4):921–950.
- Evan Sandhaus. 2008. New York Times corpus: Corpus overview. LDC catalogue entry LDC2008T19.
- The PDTB Research Group. 2008. The Penn Discourse TreeBank 2.0 Annotation Manual. Available at <http://www.seas.upenn.edu/~pdtb/>, or as part of the download of LDC2008T05.
- Bonnie Webber, Alistair Knott, and Aravind Joshi. 1999. Multiple discourse connectives in a lexicalized grammar for discourse. In *Third International Workshop on Computational Semantics*, pages 309–325, Tilburg, The Netherlands.
- Deniz Zeyrek. 2014. On the distribution of contrastive-concessive discourse connectives *ama* (but/yet) and *fakat* (but) in written Turkish. In P. Suihkonen and L.J. Whaley, editors, *On Diversity and Complexity of Languages Spoken in Europe and North and Central Asia*.